9-2010

Moore's Paradox, defective interpretation, justified belief and conscious belief

John N. WILLIAMS

Singapore Management University, johnwilliams@smu.edu.sg

DOI: https://doi.org/10.1111/j.1755-2567.2010.01073.x

Published in Theoria, September 2010, Volume 76, Issue 3, pp 221-248.
http://doi.org/10.1111/j.1755-2567.2010.01073.x

Moore's Paradox, Defective Interpretation, Justified Belief and Conscious Belief

by

JOHN N. WILLIAMS

School of Social Sciences, Singapore Management University

Abstract: In this journal, Hamid Vahid argues against three families of explanation of Moore-paradoxicality. The first is the *Wittgensteinian approach*; I assert that *p* just in case I assert that I believe that *p.* So making a Moore-paradoxical assertion involves contradictory assertions. The second is the *epistemic approach*, one committed to: if I am justified in believing that *p* then I am justified in believing that I believe that *p*. So it is impossible to have a justified omissive Moore-paradoxical belief. The third is the *conscious belief approach*, being committed to: if I consciously believe that *p* then I believe that I believe that *p*. So if I have a conscious omissive Moore-paradoxical belief, then I have contradictory second-order beliefs. In their place, Vahid argues for the *defective-interpretation approach*, broadly that charity requires us to discount the utterer of a Moore-paradoxical sentence as a speaker. I agree that the Wittgensteinian approach is unsatisfactory. But so is the defective-interpretation approach. However, there is a satisfactory version of each of the epistemic and conscious-belief approaches.

Keywords: Moore, paradox, assertion, belief, irrationality, justification, speech-acts, consciousness.

1. Introduction

MOORE OBSERVED (1942, p. 543) that to assert, "I went to the pictures last Tuesday but I do not believe that I did" would be "absurd".[1] The paradox is that this absurdity persists despite the fact that what I say about myself might be true. Moore did not notice that it is no less absurd of me to silently *believe* such a possible truth. Moore also observed (1944, p. 204) that to say, "I believe that he has gone out, but he has not" would be likewise "absurd". Unlike his first example, that has the *omissive* form:

(Om) *p* & I do not believe that *p*,

so called because it reports the omission of a specific true belief, this has the *commissive* form:

(Com) *p* & I believe that not-*p*[2]

1 "Pictures" is a rather archaic British term for "cinema".

2 Formalizing "I went to the pictures last Tuesday but I do not believe that I did" as "*p* & ~ B*p*" turns "I believe that he has gone out, but he has not" into "B*p* & ~ *p*". This commutes to "~*p* & B*p*", which may be represented as "*p* & B~*p*".

so called because it reports the commission of a specific mistake in belief.[3] What is the explanation of the absurdity of Moore-paradoxical assertions or beliefs? The question matters, because diagnoses of the pathology of Moore-paradoxicality will tell us something about rationality in speech-acts, communication, belief and action.

In a recent paper in this journal, Hamid Vahid (2008) argues against three families of explanation of Moore-paradoxicality. The first is the *Wittgensteinian approach* that Moore-paradoxical utterances are "assertorically defective" (Vahid, 2008, pp. 147, 148); I assert that *p* just in case I assert that I believe that *p*. So making a Moore-paradoxical assertion involves contradictory assertions (in the omissive case about my belief that *p*, and in the commissive case about whether *p*). The second is the *epistemic approach* that Moore-paradoxical beliefs are "epistemically defective" (Vahid, 2008, p. 153); one that is committed to: if I am justified in believing that *p* then I am justified in believing that I believe that *p*. So it is impossible to have a justified omissive Moore-paradoxical belief. An analogue of this principle delivers the same verdict in the commissive case. I propounded this approach in Williams (2004).[4] The third is the *conscious belief approach*, that conscious Moore-paradoxical belief is "doxastically defective" (Vahid, 2008, p. 149). This approach is committed to: if I consciously believe that *p* then I believe that I believe that *p*. So if I have a conscious omissive Moore-paradoxical belief, then I have contradictory second-order beliefs. Sidney Shoemaker (1988, 1995) is a proponent of this approach. I will favour the descriptor "conscious-belief approach" over Vahid's own "doxastic approach" for three reasons. Firstly, whatever is wrong with Moore-paradoxical belief is wrong epistemically, because the pathology of belief is an epistemological phenomenon. Secondly, conscious belief, not belief per se, is what occupies Vahid's discussion (2008, section 1.2). Thirdly, Vahid includes in this family of explanations the *self-falsification* approach: that it is impossible to hold an omissive *true* Moore-paradoxical belief – but this will not generalize to the commissive case, nor is it about *conscious* Moore-paradoxical belief.

Valid sees two commonalities in these three approaches. Firstly, the central principles are "ascent principles" (2008, pp. 147, 149), presumably so called because the belief-operator makes its first appearance in the right-hand side or consequent of the principle.[5] The second is that each approach identifies something

---

3 These useful terms are coined by Sorensen (1988, p. 16).

4 The only other philosopher who explains the absurdity of Moore-paradoxical beliefs in terms of epistemic notions such as justification is Claudio de Almeida, who argues that there can be no non-overridden evidence for a Moore-paradoxical belief (2001) and who uses six epistemic principles to argue that a Moore-paradoxical belief makes one incoherent (2007).

5 Vahid is not explicit on this point. It seems an exaggeration to call the three approaches "a single strategy" (2008, p. 146).

contradiction-like, but without identifying it with the content of the Moore-paradoxical assertion or belief.[6] Having objected to these sorts of explanation, Vahid argues for a Davidson-inspired *defective-interpretation approach* – that Moore-paradoxical utterances are "interpretively defective" (2008, p. 157). Broadly speaking, this is the claim that charity requires us to discount the utterer of a Moore-paradoxical sentence as a speaker.

I have my own reasons why the Wittgensteinian approach is untenable (Williams, 1998). However, while Vahid's defective-interpretation approach is unsatisfactory, there are satisfactory versions of the epistemic and conscious-belief approaches – so I will argue. Both do justice to the fact that when faced with a Moore-paradoxical speaker, we want to say that in some sense, the speaker has contradicted herself, yet we see that there is no contradiction in what she has said.

In section 2 I argue that Vahid's account is too weak, mischaracterizes the absurdity and is incomplete. I then address Anthony Brueckner's (2006) objection to my epistemic approach by modifying it in section 3. In section 4 I discuss Shoemaker's version of the conscious-belief approach, one that starts from a criticism of David Rosenthal's account of conscious belief. I argue that the Rosenthal-inspired account of the absurdity, as well as Shoemaker's, face objections. Although Vahid points out some of these, I make a number of more serious objections. This provides the basis for section 5, in which I argue for a better approach to Moore's paradox in terms of conscious belief. Then in section 6 I explain the absurdity of Moore-paradoxical assertion via that of Moore-paradoxical belief, in terms of justified belief and conscious belief. I conclude in section 7 with some remarks on unifying and developing the epistemic and consciousness approaches to a wider domain of Moore-paradoxicality.

2. Vahid's Defective-Interpretation Approach – and Three Objections

Davidson holds that a Tarskian truth theory is the appropriate form of a theory of the meaning of an object-language L: for each sentence s in it, there is a meaning-giving theorem of the form: s is true in L iff *p*, where "*p*" is the translation of s into the meta-language. When an interpreter finds that the speaker regularly assents to s under conditions she recognizes, she is entitled to take these as truth-conditions of s. For example, if we discover that a speaker regularly assents to s when and only when it is raining, we are entitled to think that s is true iff it is raining. However, one

---

6 See Heal (1994, p. 6) following Wittgenstein. In his letter to Moore, Wittgenstein notes the importance of Moore's discovery of an absurdity "which is in fact similar to a contradiction, though it isn't one" (1974, p. 177). Vahid puts this as the less nuanced claim that all three families see the absurdity as stemming from the "violation of the law of non-contradiction" (2008, p. 146).

cannot assign meanings to a speaker's utterances without knowing what the speaker believes, while one cannot identify beliefs without knowing what the speaker's utterances mean (Davidson, 1984). Accordingly, Davidson proposes a principle of charity, roughly that we interpret speakers as having true beliefs that are true by our lights, wherever it is rational for us to do so (Davidson, 1976). For example, since we normally believe that it is raining when and only when it is raining, we must also attribute that belief to a speaker of L who assents to s when and only when it is raining. On this account, which holds equally when the object-language and the meta-language are identical, we cannot take someone to be a speaker at all unless we also take her to be largely rational by our lights. Nor can we take her as a speaker unless we take her as believing what she says. As Davidson puts it:

> If we cannot find a way to interpret the utterance and other behavior of a creature as revealing a set of beliefs largely consistent and true by our standards, we have no reason to count that creature as rational, as having beliefs, or as *saying anything* (1984, p. 137; emphasis added).

Vahid applies this account to assertions of the form of:

(Om) *p* and I do not believe that *p*

as follows:

> . . . when, having assented to "it is raining", the speaker goes on to assert that she does not believe that it is raining, . . . the principle of charity is undermined . . . Moorean sentences are absurd . . . because their assertion is *interpretively* defective. It is the very status of the utterer as a speaker . . . that is put at risk . . . (Vahid, 2008, p. 160).

We may reconstruct this explanation more tersely. Since one cannot utter anything without uttering its lexical parts, in uttering (Om), I utter "*p*". But I also utter "I do not believe that *p*". If my interlocutor takes my latter utterance as expressing a truth, then she must take me as not believing my first utterance, which, by the principle of charity, means that she cannot take me to be a speaker.

The parallel explanation of the absurdity of assertions of the form:

(Com) *p* and I believe that not-*p*

is that in uttering them, I utter "*p*". But I also utter "I believe that not-*p*". If my interlocutor takes my latter utterance as expressing a truth, and also takes me as rational enough to not have contradictory beliefs, then she must take me as not believing my first utterance, which, by the principle of charity, means that she cannot take me to be a speaker (see Vahid, 2008, p. 162).[7]

---

7 In invoking the avoidance of contradictory beliefs, Vahid seems to be appealing to the "violation of the law of non-contradiction" (2008, p. 146), a move he thinks should be avoided.

This account faces three objections. The first starts with a difficulty for Davidson's account, namely that once we learn that someone is lying, and hence does not believe the content of her utterance, we must discount her as a speaker. But lies are still assertions, and we may understand the content of a lie, even once we know it is one. Indeed in judging a lie to be false, we judge it to have a meaning, for otherwise it would have no truth-value. Although a liar does not mean what she says, what she says still has a meaning.

One heroic response is to deny that lies are genuine assertions but are rather bits of play-acting (Rosenthal, 1995b, n. 15, 208). But then it would follow that I could refute your accusation that I have told you a lie by merely admitting that I was lying, for then I could not have told you anything. Davidson should reply instead that the principle of charity may be overridden by one's knowledge or justified belief that the speaker intends to deceive her interlocutor (Vahid, 2008, p. 161). But this reply will not help Vahid, for suppose that you learn that I am lying when I make a Moore-paradoxical assertion to you, and thus learn that I do not believe what I have asserted. This hardly expunges the absurdity. But since the principle of charity cannot be applied, it cannot be the explanation of the persistent absurdity. Thus Vahid's account is too weak.

Secondly, on Vahid's account we must say that someone who makes a Moore-paradoxical utterance is, as Davidson puts it, not "saying anything". But surely we do understand the content of the utterance. After all it is a central feature of the paradox that sentences of the form of (Om) or (Com) express possible truths. I may fail to believe a specific truth and I may believe a specific falsehood. It is our recognition of this that partly explains our puzzlement when confronted with a Moore-paradoxical assertion – we hear a contradiction, yet fail to find it in the content of the assertion. But if a Moore-paradoxical utterance were meaningless, it could not express a possible truth. Moreover if it were meaningless then it could not be the antecedent of a true conditional. Yet it might be true that:

If it were the case that I have the false belief that it is not raining then I would be surprised to discover that the streets are wet.

Thus Vahid's account mischaracterizes the absurdity.

Thirdly, Vahid's account is incomplete because it does nothing to explain the absurdity of Moore-paradoxical beliefs. It is no less absurd of me to *believe* a possible truth of the form of (Om) or (Com) in silence. So the absurdity of the belief, as well as the assertion, needs explanation. Thus is it strange that although Valid objects to explanations of the absurdity of Moore-paradoxical beliefs that I and Shoemaker give, he attempts no such explanation himself. In an apparent attempt to fill this *lacuna*, Vahid writes:

> Both doxastic and epistemic approaches take Moore's paradox to be essentially a paradox about belief. Thus they give priority to thought over language and meaning . . . The interpretive approach, by contrast, upholds no priority thesis. Linguistic meaning and mental content must be explained together, or not at all (Vahid, 2008, p. 162).

The priority thesis is roughly that the best strategy for an explanation of Moore-paradoxicality is to explain the absurdity of Moore-paradoxicality in language via that of Moore-paradoxical belief. But an explanation of Moore-paradoxical belief in terms of justification or conscious belief does not itself commit one to the priority thesis – although I do endorse it in section 6. More importantly, the fact that Vahid's explanation of the absurdity of Moore-paradoxical assertions makes no commitment to the priority thesis does not entail that his explanation extends to the absurdity of Moore-paradoxical belief. On Vahid's account, the putative meaninglessness of a Moore-paradoxical utterance is explained partly in terms of the absence of a mental content, namely the speaker's lack of belief in the content of her utterance. But it does not follow that a defect in mental content, in the form of Moore-paradoxical belief, has been explained in terms of linguistic meaning. In fact Vahid's account cannot explain the absurdity of a Moore-paradoxical belief held by someone who never makes the corresponding assertion, because his account must start with the *datum* of an utterance. Besides, are beliefs the sorts of things that can be interpreted? It is not clear whether this idea makes sense.[8]

3. The Epistemic Approach: the Impossibility of Justified Moore-Paradoxical Belief

In Williams (2004) I argued that Moore-paradoxical beliefs are irrational because they are impossible to justify. This argument starts with *the externalist syllogism*:

(1) All circumstances that justify me in believing that *p* are circumstances that tend to make me believe that *p*.

---

8 Vahid claims that unlike the epistemic and conscious belief approaches, his own ". . . does not depend for its success on the syntactic structure of Moorean sentences" (2008, p. 162). I agree that the syntax of (Om) or (Com) is neither necessary nor sufficient for Moore-paradoxicality. It is not necessary because my assertion or belief that God knows that I am an atheist has the essential absurdity of Moore's examples, although it does not have the form of (Om) or (Com). It is not sufficient because if eliminativism is false and so unbeknownst to himself Paul Churchland believes his consistent assertion, "There are no beliefs (any more than there are vital spirits) and I do not believe that there are no beliefs" then his belief is simply false, not absurd. But I see no reason why my versions of the epistemic and conscious-belief approaches cannot be extended to deal with these cases. Moreover, in taking the utterance of the second conjunct of (Om) or (Com) as expressing a truth, to then be compared with the utterance of the first conjunct, Vahid's interpreter needs to appeal to the syntax of conjunction.

(2) All circumstances that tend to make me believe that *p* are circumstances that justify me in believing that I believe that *p*.

So

(3) All circumstances that justify me in believing that *p* are circumstances that justify me in believing that I believe that *p*.

The syllogism appears plausible from the perspective of an externalist theory of justification, according to which justification for one's belief is a reliable or truth-conducive process of forming true beliefs, a process of which one need not be aware. Call justification from such a perspective "externalist justification". A typical case of such justification arises when what justifies me in believing that it is raining is appearing to perceive rain with my normal sensory apparatus. In such circumstances, I will tend to believe that it is raining. This supports (1). In the same circumstances, a second-order belief that I might form that I believe that it is raining, tends to be true and so is reliable. This reliability justifies that second-order belief. This establishes (2).[9] (3) explains the absurdity of believing:

(Om) *p* and I do not believe that *p*.

Suppose that there are circumstances in which I am justified in believing (Om). In these circumstances I am justified in believing its first conjunct. By (3), these are circumstances in which I am justified in believing that I believe that *p*. But in these circumstances I am also justified in believing (Om)'s second conjunct and so I am justified in believing that I do *not* believe that *p*. This is impossible, because circumstances in which I am justified in believing that something is the case are circumstances in which I am not justified in believing that it is not the case.

To explain the absurdity of believing (Com) we need:

(3′) All circumstances in which I am justified in believing that *p* are circumstances in which I am justified in believing that I do not believe that not-*p*.

(3′) follows from (3) for a subject with generally truth-conducive processes of forming beliefs. Suppose that there is a circumstance in which I have a truth-conducive process of forming the belief that *p*. By (3) I have a truth-conducive process of forming the belief that I believe that *p*. This process is *ipso facto* a truth-conducive process of forming the belief that I do not believe that not-*p*. Otherwise circumstances in which I tend to believe that *p* would be those in which

9 Vahid (2005, pp. 339–40) has made two objections to this syllogism, to which I have replied (Williams, 2006a, p. 169) in a way that Vahid himself now admits is plausible (Vahid, 2008, p. 156). The way I have just argued for (1) and (2) is a modification of the way I argued in Williams (2004, p. 350).

I tend to believe that not-*p*, with the result that my processes of forming beliefs are not generally truth-conducive after all.[10] (3′) explains the absurdity of believing:

(Com) *p* and I do believe that not-*p*.

Suppose that there are circumstances in which I am justified in believing (Com). In these circumstances I am justified in believing its first conjunct. By (3′) these circumstances are those in which I am justified in believing that I *do not* believe that not-*p*. But in these circumstances I am also justified in believing (Com)'s second conjunct and so I am justified in believing that I *do* believe that not-*p*. This is likewise impossible.

At this point I assume a distinction between being justified in believing that *p*, in other words having justification for believing that *p* – which leaves open whether one actually has the belief – and justifiably believing that *p*, where one's belief is at least partly caused by the justification for the belief, as originally explained by Roderick Firth (1978) in different terminology. Given this distinction it is definitionally true of any form of justification that:

(A) If I justifiably believe that *p* then I both believe that *p* and I am justified in believing that *p*.
$JBp \rightarrow (Bp \ \& \ B^{J}p)$

It follows that since it is impossible to be externalistically justified in believing what is Moore-paradoxical, it is also impossible actually to have an externalistically justified Moore-paradoxical belief.

Up until this point I have had externalist justification in mind. But Brueckner (2006, p. 265) objects that (1) is false once justification is conceived evidentially: a person who possesses, and thus believes, evidence *e* that justifies her in believing that *p*, might not tend to believe that *p*. I concede this point, because once justification is so conceived (1) need not be true of irrational believers.

---

10 A different way of establishing (3′) as a claim about externalist justification is via a second syllogism:

(1) All circumstances in which I am justified in believing that *p* are circumstances that tend to make me believe that *p*.

(2′) All circumstances that tend to make me believe that *p* are circumstances in which I am justified in believing that I do not believe that not-*p*.

So

(3′) All circumstances in which I am justified in believing that *p* are circumstances in which I am justified in believing that I do not believe that not-*p*.

In circumstances of standing in the rain with normal sensory apparatus, I will tend to believe that it is raining. This supports (1). In the same circumstances, a second-order belief that I do not believe that it is not raining tends to be true, otherwise I would tend to believe that it is not raining, and so my belief-forming processes would not be generally reliable. Since a second-order belief that I do not believe that it is not raining tends to be true, it is also reliable. This reliability justifies that second-order belief. This establishes (2′).

I now investigate two possible strategies in response. One strategy starts by restricting the syllogism to *rational* believers and taking justification evidentially in order to show the impossibility of being evidentially justified in believing (Om). The second strategy is based upon principles of evidential justification. I followed both strategies in Williams (2009). The first strategy now looks untenable to me while the second remains viable. On the first strategy I argued in effect (Williams, 2009, p. 492) as follows:

> First consider (1). Suppose that I am in circumstances in which I come to know or justifiably believe some *e* that would justify my belief that *p* were I to form that belief on its basis. If I am rational, then I will tend to believe that *p*. Indeed it would be irrational of me *not* to tend to believe what my possession of evidence justifies me in believing. Next, consider (2). Suppose that I am in circumstances in which I tend to believe that *p*. Since we are *also* supposing that (1) is true in order to deduce (3), these are circumstances in which I am justified in believing that *p* because I possess evidence *e*. So in sum, I am a rational agent in circumstances in which I believe *e* and in which were I to form the belief that *p* on that basis, as indeed I tend to do, my belief would be justified. This description of my circumstances is itself a justification for thinking that I believe that *p*. Of course I might not actually describe these circumstances as such. Nonetheless, if someone asks me why I think I believe that *p*, I am in a position to reply sensibly, 'That's what any rational person would think who knows something like *e*'.

I now think that this argument faces a fatal objection: if we classically think that, as long as I have a single justifier in my mental life, then I am in possession of justification for infinitely many beliefs (the infinitely many implications of whatever it is that I am initially justified in believing), it is false that I should tend, on pain of irrationality, to believe all that the available evidence justifies me in believing. So conceived evidentially, (1) is false.[11]

The second strategy is a modification of Williams (2009, pp. 494–496). It starts with a standard kind of account of evidential justification: if I am evidentially justified in believing that *p*, then for some evidence *e*, (i) I believe *e* (ii) I believe nothing that is counterevidence and (iii) if I were to believe *p* on the basis of believing *e* then I would have a justified belief that *p*, because believing *e* is part of the cause of my forming the belief that *p* – the other part including my background knowledge and beliefs. Call (iii) *the justificatory process* required to arrive at evidentially justified belief. This account is internalist in that it requires that I may be aware of my justificatory basis for believing that *p*. Now suppose that I think that I do not have the belief that *p*. For as much time as I have that second-order belief, I cannot think that the justificatory process has been completed. Thus I cannot have full awareness of my justificatory basis. This does not square with the evidentialist

---

11 I owe this objection to Claudio de Almeida and Fred Kroon. Brueckner (2009) has made other objections to this argument. Having just abandoned it, I will not consider these here.

account of justification. If it requires that I may be aware of the start of the process, by believing or knowing *e*, should it not also require that I may be aware of the end? This supports:

(B) If I believe that I do not believe that *p* then I am not justified in believing that *p*.[12]
$B{\sim}Bp \rightarrow {\sim}B^{J}p$

Next, justified belief distributes over conjunction:

(D) If I justifiably believe that (*p* & *q*) then I justifiably believe that *p* and I justifiably believe that *q*.
$JB(p \,\&\, q) \rightarrow (JBp \,\&\, JBq)$

For illustration, suppose that I justifiably believe that it is both the case that it is wet in London and that it is cold in London. Then I justifiably believe that it is wet in London. I also justifiably believe that it is cold in London. Illustrations cannot prove a principle, but the fact that the principle is apparently immune to counterexample is good enough reason to think that it is not only true but indeed constitutive of justified belief.

(A), (B) and (D) enable an *evidential argument* that it is impossible to have a justified omissive Moore-paradoxical belief:

1.	$JB(p \,\&\, {\sim}Bp)$	Suppose for *reductio ad absurdum*
2.	$JBp \,\&\, JB{\sim}Bp$	1, D
3.	JBp	2, &-elimination
4.	$JB{\sim}Bp$	2, &-elimination
5.	$B{\sim}Bp \,\&\, B^{J}{\sim}Bp$	4, A
6.	$B{\sim}Bp$	5, &-elimination

---

12 Here is a second reason. If the justificatory process goes through, then I acquire the belief that *p*, yet still believe that I do not believe that *p*. This is a failure of rationality to the extent that rational thinkers obey the principle:

(C) If I believe that I do not believe that *p* then I do not believe that *p*.
$B{\sim}Bp \rightarrow {\sim}Bp$

It might be argued that (C) is a principle of rationality because any case in which I mistakenly think that I do not believe that *p* leaves me open to epistemic criticism by the standards of introspection, given that introspection is normally an authoritative source of justification for beliefs about my mental states. But against (C) as constitutive of rationality suppose, as a case of repressed belief, that I have taken myself to not think that my ex-lover is worthy of respect. Yet faced with someone who criticizes her, I find myself defending her, and in doing so, I realize that I still believe that she is worthy of respect. It is not clear that my rationality was compromised before this realization. (I owe this example to T. Brian Mooney.) Or in another case, can I not rationally but falsely believe that I do not believe that my father was indifferent to me on the basis of an authoritative psychiatrist's misleading testimony? (I owe this example to Claudio de Almeida.) Moreover it might be said that rationality is a matter of how well I use information at my disposal, not how fallible I am in acquiring it. And perhaps all that rationality demands of introspection is reliability, not infallibility.

7.	$\sim B^J p$	6, B
8.	$Bp \,\&\, B^J p$	3, A
9.	$B^J p$	8, &-elimination
10.	$B^J p \,\&\, \sim B^J p$	9, 7, &-introduction. Contradiction
11.	$\sim JB(p \,\&\, \sim Bp)$	1, 10, *Reductio ad absurdum*

In order to prove that it is impossible to have a justified commissive Moore-paradoxical belief, we need an analogue of (B):

(B′) If I believe that I believe that not-*p* then I am not justified in believing that *p*.
$BB{\sim}p \rightarrow \sim B^J p$

The argument for (B′) is much the same as that for (B).[13] For as much time as I think that I believe that not-*p*, I cannot reasonably believe that the justificatory process has terminated in my belief that *p*. Thus I cannot have full awareness of my justificatory basis, contrary to internalism.[14] (B′) enables a second evidential argument parallel to that above:

1.	$JB(p \,\&\, B{\sim}p)$	Suppose for *reductio ad absurdum*
2.	$JBp \,\&\, JBB{\sim}p$	1, D
3.	JBp	2, &-elimination
4.	$JBB{\sim}p$	2, &-elimination
5.	$BB{\sim}p \,\&\, B^J B{\sim}p$	4, A
6.	$BB{\sim}p$	5, &-elimination
7.	$\sim B^J p$	6, B′
8.	$Bp \,\&\, B^J p$	3, A
9.	$B^J p$	8, &-elimination
10.	$B^J p \,\&\, \sim B^J p$	9, 7, &-introduction. Contradiction
11.	$\sim JB(p \,\&\, B{\sim}p)$	1, 10, *Reductio ad absurdum*

It might be objected against (B) and (B′) that at the point in time when I actually form my belief that *p* partly as a result of believing its evidential justification *e*, then as a rational thinker, I will be about to give up my belief that I do not believe that *p* – against (B) – and I will be about to give up my belief that I believe that not-*p* – against (B′).

---

13 See also de Almeida (2007, pp. 66–67) for an argument for (B′).

14 Another reason might be that if the justificatory process goes through, then I acquire the belief that *p*, yet still believe that I do not believe that *p*. This is a failure of rationality to the extent that rational thinkers obey the principle:

(E) If I believe that I believe that not-*p* then I do not believe that *p*.
$BB{\sim}p \rightarrow \sim Bp$

It might be argued that (E) is also a principle of rationality. But against (E) as constitutive of rationality, let us tweak the last example. Suppose that I have taken myself to *think* that my ex-lover is *un*worthy of respect. Yet faced with someone who criticizes her, I find myself defending her, and in doing so, I realize that I did not believe that she is unworthy of respect. It is not clear that my rationality was compromised.

I reply that although this is perfectly true, it leaves my argument untouched. For it remains true that for any period, however brief, during which I believe that I do not believe that *p* or during which I believe that I believe that not-*p*, I am not justified in believing that *p*, because during that period I cannot complete the justificatory evidential process as a rational thinker. In fact this is one explanation of why you hear absurdity when I assert (Om) or (Com): you are entitled to think that if I am sincere at the instant that I assert that I do not believe that *p* – or that I believe that not-*p* – then I cannot *yet* have succeeded in justifying the belief that *p* that I express contemporaneously.[15] Since I cannot rationally justify my belief that (*p* and I do not believe that *p*) – or my belief that (*p* and I believe that not-*p*) – the process of attempting to justify it forces me as rational thinker to give up my belief that I do not believe that *p* – or my belief that I believe that not-*p*. On the other hand if I do not revise my beliefs during the justificatory process then I am irrational to the extent that my Moore-paradoxical belief remains impossible to justify.

My evidential argument shows that it is impossible to have a Moore-paradoxical belief that is evidentially justified. My externalist syllogism shows that it is impossible to have a Moore-paradoxical belief that is externalistically justified. Assuming that these two types of justification are exhaustive, it follows that it is impossible to have a justified Moore-paradoxical belief in any sense of justification.

This result explains our puzzlement in thinking about Moore-paradoxical beliefs. We want to say that since their contents are possible truths, it should in principle be possible to justify them. But this is not so. This explains the distinctive absurdity of

---

15 There is more direct argument for the impossibility of having an evidentially justified Moore-paradoxical belief that goes as follows. For any period, however brief, during which I believe that I do not believe that *p*, I cannot, as a rational thinker, think of *e* as a basis of my belief that *p* for then I would believe that I believe that *p*, thus acquiring contradictory second-order beliefs. My inability to think of *e* as a basis of my belief that *p* is at odds with the evidentialist account of justification. This supports:

(F) If I believe that I do not believe that *p* then I do not have a justified belief that *p*.
B~B*p* → ~JB*p*

(F) enables the following proof:

1.	JB(*p* & ~B*p*)	Suppose for *reductio ad absurdum*
2.	JB*p* & JB~B*p*	1, D
3.	JB*p*	2, &-elimination
4.	JB~B*p*	2, &-elimination
5.	B~B*p* & B^{J}~B*p*	4, A
6.	B~B*p*	5, &-elimination
7.	~JB*p*	6, F
8.	JBp & ~JB*p*	3, 7, &-introduction. Contradiction
9.	~ JB(*p* & ~B*p*)	1, 8, *Reductio ad absurdum*

The commissive case can be dealt with by a parallel proof that uses:

(F′) If I believe that I do not believe that *p* then I do not have a justified belief that *p*.
BB~*p* → ~JB*p*

(F′) is supported by the fact for any period, however brief, during which I believe that I believe that not-*p*, I cannot reasonably think of *e* as a basis of my belief that *p* for then I would believe that I believe that *p*.

Moore-paradoxical belief; it is absurd to have a belief that you *cannot in principle* justify, despite the fact that it might be true. In contrast, a belief that the number of stars is odd is absurd for a different reason; unlike a Moore-paradoxical belief, the far-fetched supposition that there is justification for it is consistent.

4. The Conscious Belief Approach: Rosenthal and Shoemaker

Sidney Shoemaker (1995) gives an account of the absurdity of Moore-paradoxical belief that starts from a criticism of David Rosenthal's account of conscious belief (1997). According to Rosenthal, I am conscious of my belief that *p* just in case I have a "suitable" thought about that belief. Since my mere supposition that I have a belief would not make me aware of a belief that I really do have, the thought had best be a belief. Rosenthal observes that this second-order occurrent belief is suitable only if it represents not only the occurrence of the first-order belief, but also represents myself *as* myself in that state of belief. Consistently with this, Rosenthal propounds a higher-order principle of conscious belief:

(RP) If I consciously believe that *p* then I believe that *p* and I believe that I myself believe that *p*.
$B^{c}p \rightarrow (Bp \ \& \ B^{*}Bp)$

where "*i**" stands for "I myself", as an extension of Castañeda's (1966, 1968) quasi-indicator notation, according to which "*x* believes that *x**Φ" is read as "*x* believes that he himself Φ".[16] This *de se* element is needed. For even if I am Williams, my belief that Williams believes that *p* would not capture my awareness of my own belief. For I might not realise that I am Williams, perhaps because I am suffering from amnesia. In that case my belief that (*p* and Williams does not believe that *p*) is not absurd.

Although Rosenthal himself thinks that all that needs to be explained is the absurdity of Moore-paradoxical assertion as opposed to belief, this suggests an explanation of the absurdity of conscious Moore-paradoxical belief. This explanation needs a second principle, namely that conscious belief distributes over conjunction. It seems unobjectionable to use "aware" as a synonym of "conscious". Where *N* is a noun, whether it denotes an object such as a coin in my pocket or a mental state, surely I am conscious of having *N* just in case I am aware of having *N*. To say that I am conscious of a belief, fear, suspicion or toothache (or a coin in my pocket) is just to say that I am aware of having it. Now suppose that I become aware of my belief that

16 "*x** Φ" falls within the scope of a propositional attitude attributed to *x*, as in "*x* believes that *x** is walking with a stoop" or "*x* fears that *x** is unattractive to women". Not all uses of the reflexive pronouns can be parsed as "*x**", for example, "He cut himself accidentally" and "I disqualified myself on purpose".

it is wet and cold. Surely I then become aware of my belief that it is wet and become aware of my belief that it is cold. So it seems plausible that:

(G) If I consciously believe that (p & q) then I both consciously believe that p and I consciously believe that q.
$B^{C}(p \,\&\, q) \rightarrow (B^{C}p \,\&\, B^{C}q)$

It follows that if I consciously believe (Om) then I have contradictory beliefs. Dropping the *de se* element in (RP) for ease of exposition, the proof is:

1.	$B^{C}(p \,\&\, \sim Bp)$	Suppose
2.	$B^{C}p \,\&\, B^{C}\sim Bp$	1, G
3.	$B^{C}p$	2, &-elimination
4.	$Bp \,\&\, BBp$	3, RP
5.	BBp	4, &-elimination
6.	$B^{C}\sim Bp$	2, &-elimination
7.	$BBp \,\&\, B^{C}\sim Bp$	5, 6, &-introduction[17]

Given that a pair of contradictory beliefs is enough to explain the absurdity of conscious Moore-paradoxical belief, this proof is satisfactory as far as it goes.[18] But Shoemaker observes that it cannot explain what is wrong with Moore-paradoxical beliefs that are *not* consciously held. However, he claims that:

> . . . believing something *commits* one to believing that one believes it, in the sense that . . . if one believes something, and considers whether one does, one must, on pain of irrationality, believe that one believes it (1995, p. 214; emphasis added).

Based on this, he proposes another principle that is supposed to show that Moore-paradoxical belief is impossible, whether or not conscious, for rational subjects. This is what he calls the "self-intimation thesis":

(SI) If I believe that p then if I consider whether I believe that p, then I believe that I believe that p.
$Bp \rightarrow (B^{CON}p \rightarrow BBp)$

Shoemaker thinks that this principle will deliver the result that if I have a Moore-paradoxical belief, then I have contradictory beliefs.[19] Although he does not say as much, he needs another two principles for this to go through, namely the distribution of belief over conjunction:

(H) If I believe that (p & q) then I both believe that p, and I believe that q.
$B(p \,\&\, q) \rightarrow (Bp \,\&\, Bq)$

---

17 Vahid (2008, p. 150, n. 6) gives a similar derivation.

18 Only one belief is conscious, a point that will soon become important.

19 I will show that this result is delivered only for the omissive belief.

and the distribution of consideration over conjunction:

(I) If I consider whether I believe that (p & q) then I both consider whether I believe that p, and I consider whether I believe that q.
$B^{CON}(p \,\&\, q) \rightarrow (B^{CON}p \,\&\, B^{CON}q)$

Both of these principles are plausible. The derivation for the omissive belief goes:

1.	$B(p \,\&\, {\sim}Bp) \,\&\, B^{CON}(p \,\&\, {\sim}Bp)$	Suppose
2.	$B(p \,\&\, {\sim}Bp)$	1, &-elimination
3.	$B^{CON}(p \,\&\, {\sim}Bp)$	1, &-elimination
4.	$Bp \,\&\, B{\sim}Bp$	2, H
5.	Bp	4, &-elimination
6.	$B^{CON}p \,\&\, B^{CON}{\sim}Bp$	3, I
7.	$B^{CON}p$	6, &-elimination
8.	$B^{CON}p \rightarrow BBp$	5, SI
9.	BBp	7, 8, Modus Ponens
10.	$B{\sim}Bp$	4, &-elimination
11.	$BBp \,\&\, B{\sim}Bp$	9, 10, &-introduction

Given that a rational subject cannot have contradictory beliefs, it follows that a rational thinker cannot believe (Om).[20]

Both of these accounts are objectionable. Some objections are particular to each, but the most serious objections afflict both. The Rosenthal-inspired account fails to account for the fact that in becoming aware of having a belief, I not only become aware of that belief itself but also become aware of myself as having it. True, my second-order belief that *I* believe that it is raining represents myself. But I may be unaware of having this second-order belief. In that case my second-order belief is not a *conscious* representation of myself. So having it does not guarantee that I am aware of myself.

Rosenthal could avoid this objection by modifying his principle to:

(RP′) If I consciously believe that p then I believe that p and I *consciously* believe that I myself believe that p.
$B^{c}p \rightarrow (Bp \,\&\, B^{c}{*}Bp)$

But this would entail a vicious infinite regress. Applying (RP′) to $B^{C}p$ yields $B^{C}Bp$ which by (RP′) again yields $B^{C}BBp \ldots$ *ad infinitum*.[21] This regress is viscous because the series represents mental performances that are discrete for two reasons. Firstly, each belief is of a higher order than the one before. Secondly, on Rosenthal's account my second-order belief is a thought that occurs to me *just after* I have formed the thought that constitutes my first-order belief. So the formation of

20 Vahid (2008, p. 151) makes a similar point.

21 Strictly speaking we need to add the trivial principle that $B^{c}{*}Bp \rightarrow B^{c}Bp$.

each belief takes place after the formation of the one before. Not even the most rational thinker could complete the series that (RP′) requires. Moreover someone who has a conscious belief would have to have beliefs of as finitely large orders as we care to stipulate, and we are free to stipulate a belief of such high order that no human thinker could think the thought of its content. Such beliefs are prohibited by *Searle's Principle* (1992, pp. 155–162):

If I believe that *p* then I have the ability to think the thought that *p*.

In contrast, Shoemaker's self-intimation thesis entails no infinite regress because of the clause about hypothetical considerings.

Shoemaker's account comes with two problems of its own. The first is that the derivation of contradictory beliefs needs to start with the supposition that the Moore-paradoxical belief is actually considered. In other words, Shoemaker's account fails to explain the absurdity of Moore-paradoxical beliefs that one does *not* consider. But surely a Moore-paradoxical belief remains absurd even when its believer does not consider whether he has it. The only way out is to drop the consideration clause from the self-intimation thesis to give:

(SI′) If I believe that *p* then I believe that I believe that *p*.
$Bp \rightarrow BBp$

But now the self-intimation thesis is much less plausible, even as a claim about rational subjects. (SI′) is a principle of introspective omniscience. Surely I may have beliefs that are repressed that I do not think I have. Why should that make me irrational? True, I am ignorant of my mental states – but ignorance is not irrationality. True, the tool of introspection I have at my disposal is limited – but rationality is a matter of how well I use the tools, limited or not, that I have at my disposal. Worse still, the self-intimation thesis now entails an infinite regress of beliefs.[22]

The second problem particular to Shoemaker's account is that there are Moore-paradoxical beliefs the absurdity of which Shoemaker cannot explain. The self-intimation thesis is false in a case in which I believe that *p* but I do not believe that I believe that *p*, upon considering whether I believe that *p*. Self-deception seems to show that such a case is possible. For example, my assertion that I do not believe that women are inferior may be sincere, even after I have considered whether I believe that women are inferior, because I am blind to the way I treat women. But you may be in a better position to recognize that my boorish behavior is the manifestation of the existing belief that I sincerely deny having. In other words:

22 Vahid (2008, p. 152) makes a point similar to this last sentence.

I believe that women are inferior but I do not think that I believe they are.

Now suppose that I have a true belief that this is so. If we substitute *p* for *I believe that women are inferior*, this becomes a case in which I truly believe that (*p* but I do not believe that *p*). I now have an omissive Moore-paradoxical belief the absurdity of which cannot be explained in terms of the truth of Shoemaker's self-intimation thesis. So Shoemaker's account is incomplete.[23]

In fact neither account is complete, because neither account can explain the absurdity of *commissive* Moore-paradoxical beliefs. Supposing that I believe (Com), both accounts predict only that I believe that I both believe that *p* and believe that I believe that not-*p*. The derivation for the Rosenthal-inspired account goes:

1.	$B^{C}(p \,\&\, B{\sim}p)$	Suppose
2.	$B^{C}p \,\&\, B^{C}B{\sim}p$	1, G
3.	$B^{C}p$	2, &-elimination
4.	$Bp \,\&\, BBp$	3, RP
5.	BBp	4, &-elimination
6.	$B^{C}B{\sim}p$	2, &-elimination
7.	$BBp \,\&\, B^{C}B{\sim}p$	5, 6, &-introduction

This is not a pair of contradictory beliefs. The derivation on Shoemaker's account goes:

1.	$B(p \,\&\, B{\sim}p) \,\&\, B^{\mathbf{CON}}(p \,\&\, B{\sim}p)$	Suppose
2.	$B(p \,\&\, B{\sim}p)$	1, &-elimination
3.	$B^{\mathbf{CON}}(p \,\&\, B{\sim}p)$	1, &-elimination
4.	$Bp \,\&\, BB{\sim}p$	2, H
5.	Bp	4, &-elimination
6.	$B^{\mathbf{CON}}p \,\&\, B^{\mathbf{CON}}B{\sim}p$	3, I
7.	$B^{\mathbf{CON}}p$	6, &-elimination
8.	$B^{\mathbf{CON}}p \rightarrow BBp$	5, SI
9.	BBp	7, 8, Modus Ponens
10.	$BB{\sim}p$	4, &-elimination
11.	$BBp \,\&\, BB{\sim}p$	9, 10, &-introduction

This is not a pair of contradictory beliefs either.[24] This is related to a more serious problem, namely that the absurdity of Moore-paradoxical belief cannot be adequately explained in terms of a pair of contradictory beliefs, with the result that neither account is able even to explain the absurdity of believing (Om). The irrationality of Moore-paradoxical belief is surely severer than that of having a pair of contradictory beliefs (de Almeida, 2001, p. 43; Kriegel, 2004) for we may

23 Vahid makes this point (2008, p. 153), with the same example.

24 Nor is either belief conscious.

consistently suppose that I have contradictory beliefs because I am unaware of one or both of them. For example, a visit to a psychiatrist might unearth my long-repressed belief that my mother was an adulterer that persists in the face of my sincere adult assertion that she was not. Before the visit I held a pair of contradictory beliefs about my mother. But since I was not aware of both beliefs I was in no position to revise them. At that stage it would be harsh to judge me "absurd". A better account would diagnose the irrationality of Moore-paradoxical belief as a single belief in a self-contradiction, as when I believe that both *p* and not-*p*. It is more irrational to have a self-contradictory belief than to have a pair of contradictory beliefs. For self-contradictory beliefs are less conducive to truth than pairs of contradictory beliefs. When I have contradictory beliefs half of my beliefs are bound to be true. But none of my self-contradictory beliefs can be true. Even more absurd or irrational would be to be aware of a single self-contradictory belief while refraining from revising my beliefs. With this in mind, I propose a better account of the absurdity of Moore-paradoxical beliefs in terms of conscious beliefs.

5. The Conscious Belief Approach: A Better Version

Although belief distributes over conjunction, it does not collect over conjunction. I may be unable to think the thought of the "fat conjunction" of all of my present beliefs although I have the ability to think the thought of each of my present beliefs separately – not because I lack the relevant concepts needed to think the would-be thought, but rather because that thought is just too complex for me to think. Moreover, my beliefs appear to be innumerable. In believing that I live no more than ten miles away from London Bridge do not I also, at least dispositionally, believe that I live no more than eleven miles away from London Bridge and believe that I live no more than twelve miles away from London Bridge . . . and so on? While I have the ability to think each thought in this series, I surely do not have the ability to think the thought of their conjunction, for that would be a thought I could never finish thinking. But then Searle's principle means that I could not even have an *unconscious* belief of the conjunction of everything I now believe.

In contrast it seems that *conscious belief both distributes and collects over conjunction*:

(J) I consciously believe that (p & q) just in case I both consciously believe that p and I consciously believe that q.
$B^C(p \,\&\, q) \leftrightarrow (B^C p \,\&\, B^C q)$

This is very plausible against the background of the "synchronic unity of consciousness" (Bayne, 2008, forthcoming; Tye 2003): all the conscious states you have at a given instant are unified into a *single* encompassing state. Thus, in

consciously believing that it is wet at the same instant as consciously believing that it is cold, you are, at that instant, consciously believing that it is wet and cold. So an interesting feature of consciousness is that in becoming aware of each of a pair of contradictory beliefs, I become aware of a single belief in a self-contradiction.

But does (J) not entail a vicious regress in the same way that (RP′) does? An argument that it does is as follows. Applying (J) to $B^{C}p$ & $B^{C}q$ yields $B^{C}(p \,\&\, q)$. Applying (J) again to these three beliefs yields $B^{C}[p \,\&\, q \,\&\, (p \,\&\, q)]$. Applying (J) to these four beliefs yields $B^{C}\{p \,\&\, q \,\&\, (p \,\&\, q) \,\&\, [p \,\&\, q \,\&\, (p \,\&\, q)]\}$ . . . *ad infinitum*. But this regress does not get off the ground because the unity of consciousness really is a unity: at the instant that you become conscious of your beliefs they are agglomerated into just *one* conscious belief, with the result that there is nothing to collect. The mistake in the argument for regress is in the application of (J) to "these three beliefs". There is only one.[25]

We now need a principle in the spirit of Rosenthal's (RP) that both avoids a vicious regress of beliefs and that entails that in having a conscious belief, I am aware of myself. The principle that does the trick is:

(K) If I consciously have the *first-order belief* that p, then I both have the first-order belief that p and I consciously believe that I believe that p.
$B_1{}^{C}p \rightarrow (B_1p \,\&\, B_2{}^{C}B_1p)$ [26]

The restriction of the antecedent to first-order beliefs blocks the regress.[27] And since the consequent predicts that I am *aware* of my belief that I believe that p, this second-order belief is *conscious* representation of myself. So having it guarantees that I am aware of myself.

---

25 It might also be replied that believing that p is no different from believing that $(p \,\&\, p)$, so believing that $(p \,\&\, q)$ is no different from believing that $(p \,\&\, p \,\&\, q)$. I am not sure what to make of this reply. It might be argued that the contents of the beliefs are identical because they are logically equivalent. Against this it could be objected that a belief that a triangle is equilateral is distinct from a belief that it is equiangular despite the fact that their contents are logically equivalent, because I could have the former without being able to have the latter if I have the concept of equal length yet have no clue what angles are – as dictated by Searle's principle. It might even be claimed that I could believe that p without being able to believe that $(p \,\&\, p)$ if I lack the concept of conjunction.

26 Strictly speaking, this should be:

> If I consciously have the *first-order belief* that p, then I both have the first-order belief that p and I consciously believe that *I myself* believe that p.
> $B_1{}^{C}p \rightarrow (B_1p \,\&\, B_2{}^{C*}B_1p)$.

For ease of exposition, I drop the *de se* element in what follows. Nothing turns upon this.

27 This is an improvement upon my earlier suggestion (Williams, 2006b, p. 402) of the principle:

> If I consciously believe that p then I both believe that p and I consciously believe that I believe that p.
> $B^{C}p \rightarrow (Bp \,\&\, B^{C}Bp)$

where the principle comes with the restriction that it may only be applied a finite number of times, in order to avoid infinite regress. It might be objected that this restriction is *ad hoc*.

I need only these two principles to explain the absurdity of Moore-paradoxical belief. If I consciously have the omissive belief that (p & I do not believe that p) then I consciously believe that I *both do and do not* believe that p. The proof of this is:

1.	$B_2^C(p \ \& \ \sim B_1 p)$	Suppose conscious omissive belief
2.	$B_1^C p \ \& \ B_2^C \sim B_1 p$	1, J
3.	$B_1^C p$	2, &-elimination
4.	$B_2^C \sim B_1 p$	2, &-elimination
5.	$B_1 p \ \& \ B_2^C B_1 p$	3, K
6.	$B_2^C B_1 p$	5, &-elimination
7.	$B_2^C B_1 p \ \& \ B_2^C \sim B_1 p$	6, 4, &-introduction
8.	$B_2^C(B_1 p \ \& \ \sim B_1 p)$	6, J

This result provides a pathology of the Moore-paradoxical believer: in becoming conscious of my omissive belief, I become aware that I believe a self-contradiction. Unless I change my beliefs, I am mad. A different pathology emerges for the commissive belief: if I consciously have the commissive belief that (p & I believe that not-p) then I consciously believe that I have a pair of contradictory beliefs. The proof of this is:

1.	$B_2^C(p \ \& \ B_1 \sim p)$	Suppose
2.	$B_1^C p \ \& \ B_2^C B_1 \sim p$	1, J
3.	$B_1^C p$	2, &-elimination
4.	$B_2^C B_1 \sim p$	2, &-elimination
5.	$B_1 p \ \& \ B_2^C B_1 p$	3, K
6.	$B_2^C B_1 p$	5, &-elimination
7.	$B_2^C B_1 p \ \& \ B_2^C B_1 \sim p$	6, 4, &-introduction
8.	$B_2^C(B_1 p \ \& \ B_1 \sim p)$	6, J

In other words, in becoming conscious of my commissive belief, I become aware that I have contradictory beliefs. Unless I revise my beliefs, I am damaged goods. We might expect a difference in pathology given the difference between a specific instance of ignorance (in the omissive case) and a specific mistake in belief (in the commissive case). The epistemic approach cannot explain this difference since it diagnoses the pathology as the impossibility of justification in both cases.

The occurrence of conscious belief in (Om) or (Com) amounts to a recognition of irrationality. The irrationality is not a good thing, but the recognition of it is, because it amounts to the capacity for belief-revision. If that epiphany leads me to no epistemic revision, then I am indeed irrational. For then I realize that I have contradictory or self-contradictory beliefs, yet continue to have them. Such a case might occur when exasperated by a particularly obtuse psychiatrist who keeps reassuring me that my belief that I am the victim of persecution is just a delusion, I remark, "Look here, I jolly well know that people aren't persecuting me, but I just can't help believing that they are!"

There remains a question to be answered: how is the absurdity of *unconscious* Moore-paradoxical belief to be explained?[28] The answer lies in the fact that having a Moore-paradoxical belief falsifies it – or in the commissive case, falsifies it on pain of having contradictory beliefs. First take the omissive belief. If I believe that (*p* & I do not believe that *p*), then since belief distributes over conjunction, I believe that *p*. But then what I believe is false, since its second conjunct is false. Although my belief is not a belief in a necessary falsehood, it is *self-falsifying*. In other words, although what I believe might be true of me and although I might believe it, it cannot be true of me *if* I believe it. In contrast, I *can* have a true commissive belief. For if I believe that (*p* & I believe that not-*p*) then since belief distributes over conjunction, again I believe that *p*, which contradicts the second conjunct of what I believe. So unless (Com) is false, I have contradictory beliefs about whether *p*.

Suppose however that I have the Moore-paradoxical belief but am not conscious of it. It seems that I am still irrational in some sense. My belief contains a structural flaw, even if I do not see it. It is tempting to diagnose this lesser irrationality as the fact that having the belief is what leads me inescapably to falsehood. However, there is reason to think that this is an overgeneralization. For suppose, as a case of epistemic modesty, that:

(α) I believe that (β) at least one of my beliefs is false.

Necessarily, either my second-order belief reported by (α) is true or it is false. If it is true then it is true. If it is false, then (β) is false. In that case all my beliefs are true and therefore my second-order belief reported by (α) is also true. So necessarily, my second-order belief reported by (α) is true. That belief is *self-verifying*. Once I believe (β), (β) must be true.[29] So if I believe (β) then I must have at least one false belief. In other words, if I believe (β) then I have beliefs that are *inconsistent*, in the sense that at least one of them must be false. But it may be irrational of me *not* to believe (β) given inductive evidence about beliefs I have had in the past that turned

28 It is important to note that "unconscious" is not meant as a synonym of "dispositional". Essentially, a dispositional belief, such as your belief that zebras do not wear mink coats, is one that you will form when prompted. No belief, while only dispositional, has been formed. Since there is no belief to be aware of, dispositional beliefs are unconscious. But not all unconscious beliefs are dispositional. One may be unaware of beliefs that one has just formed, which one might call "occurrent" or that one has had for some time. For example, you may have formed rapidly changing perceptual or repressed beliefs of which you are unaware.

29 Prior (1971, p. 85) makes a similar point about the preface paradox in terms of assertion rather than belief.

out to be false.[30] In that case, rationality demands that I have inconsistent beliefs.[31] My belief in (β) leads me inescapably to falsehood, yet I am not thereby irrational. This result also shows that Mitchell Green is mistaken to diagnose the absurdity of Moore-paradoxical belief as "a severe violation of theoretical rationality" (2007, p. 191) just because it is a case in which "my system of beliefs is guaranteed to put me in error no matter how the world happens to be, and in a way that I could in principle discern with no empirical investigation" (2007, p. 191).[32] For we have just discerned with no empirical investigation that my belief in (β) is guaranteed to put me in error no matter how the world happens to be.

The truth of my belief that I have at least one false belief does not entail beliefs that contradict each other. Clearly I need not believe that all of my beliefs are true. We have already seen that I cannot believe the "fat conjunction" of all my beliefs. Even if I did believe it, it would not contradict my belief that I have at least one false belief, unless the fat conjunction includes the final conjunct "and these are all the beliefs I have". We are unable to believe this extra conjunct. We are in no position to list the innumerably many beliefs we have, many of which we are unaware of having.

This explains why a commitment to the necessity of at least one false belief is benign. Inconsistency in my beliefs need not undermine my justification in the way my self-contradictory or contradictory beliefs do. Justification for my belief that I have at least one false belief, such as the fact that I have held false beliefs in the past, need not count against any particular one of the vast number of other beliefs I now have. Nor will justification for any particular one of these other beliefs count in favour of my infallibility.

Faced with this difficulty of epistemic modesty we must content ourselves with diagnosing the irrationality of unconscious Moore-paradoxical belief as the fact that having a Moore-paradoxical belief falsifies it – in the commissive case, on pain

---

30 Doris Olin (2003, pp. 68–69) objects that induction from past error is illegitimate. Suppose I arrive at my beliefs by reading tea-leaves. That method has led to error, yet the inference that new beliefs so formed are likely to be in error is illegitimate because the set of beliefs that are the negations of the former beliefs would not be less likely to be in error. Richmond Campbell (2004, p. 311) points out that that is because the connection between tea-leaf reading and truth is random. Given that, the likelihood of error is great in either set, but this fact in no way undermines the original inference.

31 This case is importantly different from the simple Liar sentence: L) L is false. The truth of L) entails its falsehood and its falsehood entails its truth. In contrast, although (α) refers to itself, its truth does not entail its falsehood nor does its falsehood entail its truth; and likewise for (β). This argument for the possibility of rational inconsistent beliefs is far simpler than preface-paradox arguments to which Doris Olin objects (2003, pp. 65–70) and is not vulnerable to these.

32 This last condition is needed because Jane is not absurd in believing that Hesperus is shining but Phosphorous is not, if she needs empirical investigation to discover that Hesperus is Phosphorous (Green, 2007, p. 192). Her system of beliefs is guaranteed to put her in error no matter how the world happens to be, but not in a way that she could in principle discern with no empirical investigation.

of having contradictory beliefs. In either case I am in an epistemically bad position; a self-falsifying belief is as useless as a guide away from falsehood as a pair of contradictory beliefs. I am responsible in a special way for being in this position, because what puts me there is not the content of my belief but my forming it.[33] I have shot myself in the foot. Nonetheless, if I am unaware of having the belief then I can hardly be expected to see that this is so. Moreover, I can hardly be blamed for not revising beliefs of which I am unaware. I submit that this result coheres nicely with our intuition that any irrationality in unconscious Moore-paradoxical belief is far milder than the severe irrationality of conscious Moore-paradoxical belief.

6. The Absurdity of Moore-Paradoxical Assertion

Let us now take a closer look at the priority thesis – roughly that the best strategy for an explanation of Moore-paradoxicality is to explain the absurdity of Moore-paradoxicality in language via that of Moore-paradoxical belief. Its leading exponent, Shoemaker, asserts the restriction that:

What can be coherently believed constrains what can be coherently asserted

but adds that the converse does not hold. Since "coherently" might mean "consistently", "appropriately", "intelligibly" or "rationally", the restriction is best elucidated using Moore's own term "absurdly", by which he seems to mean "irrationally, either in theory or practice". Then Shoemaker's restriction becomes:

If my belief that *p* is absurd then so is my assertion that *p*, but not conversely.

I can offer no knock-down argument for this principle. But it is plausible to the extent that it is apparently immune to counterexample. The failure of its converse is shown by my assertion:

I am asserting nothing now.

Such an assertion would be absurd. Although the content of my assertion might be true and although I might assert it, it cannot be true *if I* assert it. As a reasonable speaker I would recognize that the assertion is self-falsifying in this way and so would not make it. But it need not be at all absurd of me to *believe* that I am asserting nothing now. I might be meditating in church.

---

33 In contrast, when I believe a necessary falsehood, the content of my belief and my believing it conspire to put me in a bad epistemic position.

Shoemaker's formulation of the priority thesis is that because sincere assertions involve believing their content, once an explanation of the impossibility of believing them is at hand, "an explanation of why one cannot assert a Moore-paradoxical sentence will come along for free" (Shoemaker, 1995, p. 213).

It is helpful to remember that Shoemaker is talking of rational agents. I agree that we should require an explanation of why Moore-paradoxical beliefs cannot be rational. Nor can it be rational to make the corresponding assertion. But there are two problems with the way Shoemaker makes his point. In the first place, applying the priority thesis only to assertions that are sincere will not explain the absurdity of Moore-paradoxical lies, as we saw in discussing Vahid's account. Secondly, it is an exaggeration to say that the explanation of the irrationality of the assertion comes "free" with that of the corresponding belief. For once we have explained why the belief is irrational we will also have to explain why this makes the corresponding assertion irrational.

A popular explanation of this is that an assertor is irrational if she "expresses" an irrational belief. But that in turn means that the notion of expressing belief will need elucidation – a need that is not always met in the literature.[34] And it will also need to be explained why, on this elucidation, expressing an irrational belief is itself irrational.

Both my approaches satisfy both needs. On the epistemic approach, my Moore-paradoxical assertion is absurd because in making such an assertion, I express a belief that cannot in principle be justified. I assert to you that *p* just in case I ostensibly express my belief that *p* to you with the intention of changing your beliefs in a relevant way. The reference to ostensible expression accommodates lies, which are surely genuine assertions. The change in your beliefs that I intend to bring about is relevant in the sense that the proposition I assert forms the core of the description of that change. There are types of changes, corresponding to types of assertion. For example, in lying to you that *p* I intend to get you to believe falsely that *p*, and in letting you know that *p* I intend to impart to you my knowledge that *p*. When I protest that I am innocent of a crime yet know that I cannot convince you of my innocence, I might sensibly aim to make you think that I am convinced of my own innocence. I ostensibly express my belief that *p* to you just in case I behave in a way that intentionally offers you reason to think that I believe that *p*.[35] When my assertion is Moore-paradoxical, what I offer you is a reason to think that I have a belief that I cannot possibly justify. Since such beliefs are irrational, I have offered

---

34 No elucidation of the term is given by Wittgenstein (1980, §472), Heal (1994, p. 22), Hájek and Stoljar (2001) or Rosenthal (1995b, pp. 197, 199, and 1995a, pp. 317–319), all of whom follow the "expressivist" approach described above.

35 In Williams (2007, pp. 154–155) I show that more exotic types of assertion ("winding you up", double-bluffs and "stonewalling lies") that appear problematic for this account do not in fact threaten it.

you a license to think that I am irrational. Given my charitable presumption that you will charitably avoid judging me irrational, I am in a position to see that you will not accept that I believe what I assert. So my assertion requires trying to achieve something I should see will not succeed and thus makes me irrational in practice.

An exception occurs when my aim in making the assertion is to make you think that I am irrational as part of a rational attempt to pretend to be insane.[36] Any absurd assertion might serve this purpose, Moore-paradoxical or not. Then my assertion will not count as any of the speech-acts above, although lying is the closest because the pretence involves deception. In this case I intend that my charitable presumption – that you will charitably avoid judging me irrational – will be defeated. Any satisfactory explanation of Moore-paradoxical assertion will allow for this exception.

On the conscious belief approach, the explanation is different. You have no reason to believe what I assert to you unless you think that *I believe* my own assertion. In other words, believing my assertion requires that you believe *me*, in the sense that you believe that I am sincerely telling the truth. Now suppose that you believe that I am sincerely telling the truth when I assert to you that (*p* and I do not believe that *p*). You must think that I believe that *p* (in virtue of now thinking me sincere) and you must, in the same instant, believe that I do not believe that *p* (in virtue of now thinking me to be telling the truth). These are conscious thoughts that you form as I make my assertion to you. Since your conscious beliefs collect over conjunction, you consciously believe that I both do and do not believe that *p*. On my charitable presumption that you are not irrational, I am in a position to see that you cannot believe me.

Parallel reasoning applies to the commissive assertion. Suppose that you come to believe me when I assert to you that (*p* and I believe that not-*p*). You must now think that I believe that *p* (in virtue of now thinking me to be sincere) and you must now, in the same instant, believe that I believe that not-*p* (in virtue of now thinking me to tell the truth). Since these two beliefs you have just acquired are conscious thoughts, you now consciously believe that *I* have contradictory beliefs about whether *p*. On my charitable presumption that you will charitably avoid judging me irrational, I am again in a position to see that you cannot believe me.

There is another explanation as well. When I make an assertion to you, you are normally entitled to think that I am not mumbling in my sleep. In that case you have no reason to believe what I assert to you unless you think that I *consciously* believe my own assertion. When my assertion is Moore-paradoxical, this entitles you to think that I am aware of believing a self-contradiction – or that I am aware of having a pair of contradictory beliefs – as just discussed in section 5. Since I should see that you would be charitable enough to try to avoid judging me irrational in this

36 I owe this point to Claudio de Almeida.

way, I should once again see that you could not believe me. Since this is part of my aim, making the assertion makes me irrational in practice.

7. Concluding Remarks

There appear to be other propositions that are absurd to believe, and therefore to assert, in the same way as (Om) or (Com). Plausible candidates include:

I have no beliefs

and Sorensen's (1988, p. 17):

Although you do not agree with me about anything, you are always right[37]
God knows that we are atheists

as well as others that might be related, such as:

I believe that it is raining but I have no justification for believing that it is raining
It is raining but I do not know that it is raining.

There are also "self-referential" candidates such as:

I believe that this sentence is false

as well as problem cases, as when Paul Churchland makes the consistent assertion that:

There are no beliefs (any more than there are vital spirits) but I do not believe that there are no beliefs.

How will the epistemic and consciousness approaches handle these cases? Will they co-operate in handling them? More generally, how do we go about producing a full or partial analysis of Moore-paradoxicality in the first place, so that we may judge which cases share the essential absurdity of Moore's own examples? If I am on the right lines so far, these are questions I should answer. But that seems best left as a separate task.

Acknowledgements

This work was supported by a project (08-C242-SMU-009) from the Office of Research, Singapore Management University. I am grateful to Fred Kroon and Claudio de Almeida for tough incisive criticism.

37 This might be put more colloquially as "Although you think all my opinions are mistaken, you are always right".

References

BRUECKNER, A. (2006) "Justification and Moore's Paradox." *Analysis*, 66(3): 264–266.

BRUECKNER, A. (2009) "More on Justification and Moore's Paradox." *Analysis*, 69(3): 497–499.

BAYNE, T. (2008) "The Unity of Consciousness and the Split-Brain Syndrome." *Journal of Philosophy*, 105(6): 277–300.

BAYNE, T. (forthcoming) *The Unity of Consciousness*. Oxford: Oxford University Press.

CAMPBELL, R. (2004) "Review of Doris Olin: Paradox." *University of Toronto Quarterly*, 74(1): 311–312.

CASTAÑEDA, H. N. (1966) " 'He': A Study in the Logic of Self-Consciousness." *Ratio*, 8(1): 130–157.

CASTAÑEDA, H. N. (1968) "On the Logic of Attribution of Self-Knowledge to Others." *Journal of Philosophy*, 65(14): 439–456.

DAVIDSON, D. (1976) "Reply to Foster." Repr. in D. Davidson (1984), *Inquiries into Truth and Interpretation*, pp. 171–179. Oxford: Oxford University Press.

DAVIDSON, D. (1984) *Inquiries into Truth and Interpretation*. Oxford: Oxford University Press.

DE ALMEIDA, C. (2001) "What Moore's Paradox is About." *Philosophy and Phenomenological Research*, 62(1): 33–58.

DE ALMEIDA, C. (2007) "Moorean Absurdity: An Epistemological Analysis." In M. S. Green and J. N. Williams (eds), *Moore's Paradox: New Essays on Belief, Rationality, and the First Person*, pp. 53–75. Oxford: Oxford University Press.

FIRTH, R. (1978) "Are Epistemic Concepts Reducible to Ethical Concepts?" In A. Goldmand and J. Kim (eds), *Values and Morals*, pp. 215–229. Dordrecht: D. Reidel Publishing Co.

GREEN, M. S. (2007) "Moorean Absurdity and Showing What's Within." In M. S. Green and J. N. Williams (eds), *Moore's Paradox: New Essays on Belief, Rationality, and the First Person*, pp. 189–214. Oxford: Oxford University Press.

HEAL, J. (1994) "Moore's Paradox: A Wittgensteinian Approach." *Mind*, 103(409): 5–24.

HÁJEK, A. and STOLJAR, D. (2001) "Crimmins, Gonzales, and Moore." *Analysis*, 61(3): 208–213.

KRIEGEL, U. (2004) "Moore's Paradox and the Structure of Conscious Belief." *Erkenntnis*, 61(1): 99–121.

MOORE, G. E. (1942) "A Reply to My Critics." In P. Schilpp (ed.), *The Philosophy of G.E. Moore*, pp. 535–667. Evanston: Northwestern University Press.

MOORE, G. E. (1944) "Russell's Theory of Descriptions." In P. Schilpp (ed.), *The Philosophy of Bertrand Russell*, pp. 175–225. Evanston: Northwestern University Press.

OLIN, D. (2003) *Paradox*. Montreal: McGill-Queen's University Press.

PRIOR, A. N. (1971) *Objects of Thought*, eds. P. T. Geach and A. J. P. Kenny. Oxford: Clarendon Press.

ROSENTHAL, D. M. (1995a) "Moore's Paradox and Consciousness." *Philosophical Perspectives*, 9: 313–333.

ROSENTHAL, D. M. (1995b) "Self-Knowledge and Moore's Paradox." *Philosophical Studies*, 77(2/3): 195–209.

ROSENTHAL, D. M. (1997) "A Theory of Consciousness." In N. J. Block, O. Flanagan and G. Güzeldere (eds), *The Nature of Consciousness: Philosophical Debates*, pp. 729–753. Cambridge, MA: MIT Press and Bradford Books.

SEARLE, J. (1992) *The Rediscovery of the Mind*. Cambridge, MA: MIT Press.
SHOEMAKER, S. (1995) "Moore's Paradox and Self-Knowledge." *Philosophical Studies*, 77(2/3): 211–228.
SHOEMAKER, S. (1988) "On Knowing One's Own Mind." *Philosophical Perspectives*, 2: 183–209.
SORENSEN, R. A. (1988) *Blindspots*. Clarendon Press: Oxford.
TYE, M. (2003) *Consciousness and Persons: Unity and Identity*. Cambridge, MA: MIT Press.
VAHID, H. (2005) "Moore's Paradox and Evans's Principle: A Reply to Williams." *Analysis*, 65(4): 337–341.
VAHID, H. (2008) "Radical Interpretation and Moore's Paradox." *Theoria*, 74(2): 146–163.
WILLIAMS, J. N. (1998) "Wittgensteinian Accounts of Moorean Absurdity." *Philosophical Studies*, 92(3): 283–306.
WILLIAMS, J. N. (2004) "Moore's Paradoxes, Evans's Principle and Self-Knowledge." *Analysis*, 64(4): 348–353.
WILLIAMS, J. N. (2006a) "In Defence of an Argument for Evans's Principle: a Rejoinder to Vahid." *Analysis*, 66(2): 167–170.
WILLIAMS, J. N. (2006b) "Moore's Paradoxes and Conscious Belief." *Philosophical Studies*, 127(3), 383–414.
WILLIAMS, J. N. (2007) "Moore's Paradoxes and Iterated Belief." *Journal of Philosophical Studies*, 32: 145–168.
WILLIAMS, J. N. (2009) "Justifying Circumstances and Moore-Paradoxical Beliefs: a Response to Brueckner." *Analysis*, 69(3): 490–496.
WITTGENSTEIN, L. (1974) *Letters to Russell, Keynes and Moore*, eds. G. H. von Wright and B. F. McGuinness. Oxford: Blackwell.
WITTGENSTEIN, L. (1980) *Remarks on the Philosophy of Psychology*, Vol. 1, eds. G. Anscombe and G. von Wright. Chicago: University of Chicago Press.